Claims

We claim:

1.   A system for constructing a relational database with associated physical structures for storing XML data wherein the XML data has a corresponding XSD schema and a workload comprising queries that have been executed on the XML data comprising:

a mapping transformation enumerator that examines queries in the workload and generates candidate mapping transformations based on the queries wherein each candidate mapping transformation can be used to transform a default mapping to a candidate mapping from the XSD schema to relational database schema and wherein each candidate transformation is added to a candidate pool;

a physical design tool that associates a set of physical design structures with a candidate mapping based on queries in the workload; and

a design tuner that searches candidate mappings and associated physical design structures and selects a preferred mapping and associated physical design structure.

2.   The system of claim 1 comprising a default mapping construction tool that performs a hybrid inlining mapping on the XSD schema to construct the default mapping.

3.   The system of claim 1 wherein the examined query accesses a single child of choice node in the XSD schema and wherein the mapping transformation enumerator enumerates a union distribution transformation on the accessed node.

4.   The system of claim 1 wherein the examined query accesses an optional node in the XSD schema and wherein the mapping transformation enumerator enumerates an implicit union distribution transformation on the accessed node.

27

5.   The system of claim 1 wherein the examined query refers to a set-valued element in the XSD schema and wherein the mapping transformation enumerator enumerates a pull-up transformation for the set-valued element.

6.   The system of claim 5 wherein the mapping transformation enumerator generates a pull-up for the set valued element if the set-valued element has a maximum cardinality of less than a threshold cardinality.

7.   The system of claim 5 wherein the mapping transformation enumerator generates a pull-up for the set valued element if more than a threshold number of the set-valued elements have a cardinality of less than a threshold cardinality the set-valued element has a maximum cardinality.

8.   The system of claim 1 further comprising a candidate transformation merger that merges candidate transformations in the candidate pool to form a merged candidate transformation.

9.   The system of claim 8 wherein the candidate merger merges candidates having optional nodes that were created using an implicit union distribution and wherein the candidates are merged on their optional nodes.

10.   The system of claim 1 wherein the design tuner enumerates mappings generated from the default mapping by applying a sequence of candidate transformations to the default mapping.

11.   The system of claim 1 wherein the design tuner selects a preferred mapping and associated physical design structures by calling a cost estimator to estimate a cost to execute queries in the workload on a relational database resulting from the

mapping in the presence of the associated physical design structures and selects the mapping and physical design structures with the lowest cost.

12.   The system of claim 11 wherein the design tuner estimates cost by deriving an estimated cost from a known cost for another mapping.

13.   The system of claim 1 wherein the physical design tool associates physical design structures with a candidate mapping by building a relational database schema using the candidate mapping and selecting physical design structures based on relational database queries corresponding to queries in the workload.

14.   The system of claim 13 comprising a set of statistical information about the XML database that is accessed by the physical design tool to select physical design structures to associate with a candidate mapping.

15.   The system of claim 13 wherein the set of statistical information is compiled by populating a relational database created by applying the default mapping to the XSD schema with sample data from the XML data.

16. A method that constructs a relational database with associated physical structures for storing XML data having an associated XSD schema and a workload made up of queries that have been executed on the XML data comprising:

examining queries in the workload to generate candidate mapping transformations based on the queries wherein a candidate mapping transformation can be used to transform a default mapping to a candidate mapping from the XSD schema to a relational database schema and wherein each candidate transformation is added to a candidate pool;

associating a set of physical database design structures with each candidate transformation based on the workload; and

searching candidate mappings and associated physical database design structures and selecting a preferred mapping and associated physical design structure.

17. The method of claim 16 further comprising selecting a mapping and associated physical design structures from among the enumerated mappings based on the performance of a relational database implementing the mapping and associated physical design structure with respect to the workload.

18. The method of claim 16 further comprising constructing the default mapping by transforming the XSD schema using a hybrid inlining mapping.

19. The method of claim 16 wherein if the examined query accesses a single child of a choice node in the XSD schema, a mapping transformation is selected that transforms the given mapping using a union distribution on the accessed node.

20. The method of claim 16 wherein if the examined query accesses an optional node in the XSD schema a mapping transformation is selected that transforms the given mapping using an implicit union distribution on the accessed node.

21. The method of claim 16 wherein if the examined query refers to a set-valued element in the XSD schema a mapping transformation is selected that transforms the given mapping by generating a pull-up for the set-valued element.

22. The method of claim 21 wherein the mapping transformation that generates a pull-up for the set valued element is selected if the set-valued element has a maximum cardinality of less than a threshold cardinality.

23. The method of claim 21 wherein the mapping transformation that generates a pull-up for the set valued element is selected if more than a threshold number of the set-valued elements have a cardinality of less than a threshold cardinality.

24. The method of claim 16 comprising merging candidate mapping transformations to form merged candidate mapping transformations.

25. The method of claim 24 wherein candidates that have optional nodes that were created using an implicit union distribution are merged and wherein the candidates are merged on their optional nodes.

26. The method of claim 16 wherein the preferred mapping is selected by estimating a cost to execute queries in the workload in a relational database implementing the mapping and the associated physical design structures and selecting the mapping and physical design structures with the lowest cost.

27. The method of claim 26 wherein the cost is estimated by deriving an estimated cost from a known cost for another mapping.

28.   A computer readable medium having computer executable instructions stored thereon for performing the method of claim 16.

29. Computer readable media having computer-executable instructions stored thereon for constructing a relational database with associated physical structures for storing XML data having a corresponding XSD schema and a workload comprising queries that have been executed on the XML data, the instructions comprising:

examining a queries in the workload and generating candidate mapping transformations based on queries in the workload wherein a candidate mapping transformation can be used to generate mappings from XSD to relational database schema and wherein each candidate transformation is added to a candidate pool comprising a set of candidate transformations; and

searching mappings generated from transforming a default mapping using candidate transformations in the candidate pool together with physical design structures associated with those mappings and selecting a preferred mapping and associated physical design structures.

30. The computer readable media of claim 29 wherein the instructions comprise transforming the XSD schema to a relational database schema using a default mapping protocol so that the default mapping is transformed based on the query.

31. The computer readable media of claim 29 wherein the examined query accesses a single child of a choice node in the XSD schema and wherein the instructions comprise transforming the given mapping using a union distribution on the accessed node.

32. The computer readable media of claim 29 wherein the examined query accesses an optional node in the XSD schema and wherein instructions comprise

transforming the given mapping using an implicit union distribution on the accessed node.

33. The computer readable media of claim 29 wherein the examined query refers to a set-valued element in the XSD schema and wherein the instructions comprise transforming the given mapping by generating a pull-up for the set-valued element.

34. The computer readable media of claim 29 wherein the instructions comprise merging candidate mapping transformations in the candidate pool to form a merged candidate mapping transformation.

35. The computer readable media of claim 34 wherein candidates having optional nodes that were created using an implicit union distribution are merged and wherein the candidates are merged on their optional nodes. .

36. The computer readable media of claim 29 wherein the instructions comprise enumerating mappings generated from the default mapping by applying sequence of transformations in the candidate pool to the default mapping, associating physical design structures with each such mapping, and selecting a preferred mapping based on the mapping and associated physical design structures.

37. The computer readable media of claim 36 wherein instruction comprise selecting a preferred mapping with its associated physical design structures by estimating a cost to execute queries in the workload in a relational database implementing the mapping and physical design structures and selecting the mapping and associated physical design structures with the lowest cost.

38. The computer readable media of claim 37 wherein the cost is estimated by deriving an estimated cost from a known cost for another mapping.

39.   A method for producing a relational database schema for population by XML data having a corresponding XSD schema and a workload comprising queries that have been executed on the XML data comprising:

transforming the XSD schema to a default mapping that maps the XSD schema to a relational database schema using a default mapping protocol;

generating a candidate mapping by transforming the default mapping based on at least one query in the workload; and

transforming the XSD schema to a relational database schema using a selected candidate mapping.

40.   The method of claim 39 wherein if a query accesses a single child of choice node in the XSD schema, a union distribution transformation is performed on the accessed child of choice node in the default mapping.

41.   The method of claim 39 wherein if the a query accesses an optional node in the XSD schema, an implicit union distribution transformation is performed on the accessed optional node in the default mapping.

42.   The method of claim 39 wherein if a query refers to a set-valued element in the XSD schema, a pull-up transformation is performed for the set-valued element in the default mapping.